PoseShield: Neural Collision Fields for Human Self-Collision Resolution

Zhengyuan Li[1], Zeyun Deng[1], Yifan Shen[3], Liangyan Gui[3], Miaolan Xie[1], Joseph Campbell[1], Xifeng Gao[2], Kui Wu[2], Zherong Pan[2], and Aniket Bera[1]

[1] Purdue University, West Lafayette IN 47907, USA
[2] LightSpeed Studios, Bellevue WA 98004, USA
[3] University of Illinois Urbana-Champaign, Champaign IL 61820, USA

Abstract. Self-collision remains a persistent challenge in SMPL-based human pose estimation and motion generation. Under extreme articulations or stochastic motion synthesis, generated meshes frequently exhibit self-penetrations, leading to physically implausible results. We propose PoseShield, a neural collision constraint defined directly in SMPL pose space. We formulate collision correction as a constrained optimization problem and connect the learned constraint with the Eikonal equation. Enforcing Eikonal regularization ensures non-vanishing gradients near the collision boundary, improving numerical stability and robustness of the optimization process. Unlike prior methods that operate in the mesh space or rely on heuristic penalties, our approach operates directly in the low-dimensional space of human poses and is theoretically grounded. The same learned constraint extends to human motion sequences, providing a generator-agnostic post-hoc collision corrector without retraining the underlying motion model. Experiments on a newly constructed SMPL pose benchmark show that our method achieves a 95.8% success rate and outperforms state-of-the-art baselines.

1 Introduction

Parametric human body models such as SMPL [33], SMPL+H [46], and SMPL-X [44] have become the standard geometric representation in human pose estimation [4, 7, 12, 23, 26, 64, 65] and motion generation [28, 32, 55, 63]. Thanks to their explicit mesh topology and low-dimensional pose parameterization, these models enable efficient optimization and tight integration with learning-based pipelines. Despite their widespread adoption, *self-collision* remains a persistent challenge.

Self-collisions arise across diverse SMPL-based applications, ranging from motion-capture-based reconstruction to motion synthesis. For reconstruction, COAP [36] shows that poses from datasets such as PROX [18] may contain body self-intersections. For motion synthesis, recent self-collision-aware generation work [19] further analyzes that modern motion synthesis methods [16,56] are prone to producing motions with non-negligible self-intersections under stochastic sampling. Such artifacts degrade physical plausibility. Consequently, there

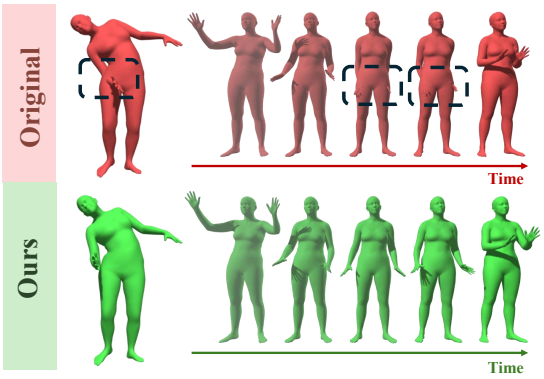


Fig. 1: PoseShield provides a robust collision constraint for SMPL-based bodies. While effective for individual human poses, it naturally extends to temporally consistent motion generation.

is a pressing need for a *reliable, post-hoc self-collision resolution approach*. An ideal approach must decouple collision handling from specific generative priors to act as a universal refinement module, ensuring geometric consistency across diverse SMPL-based scenarios without relying on original generative conditions like textual prompts or reference images.

Classical collision-handling methods have been extensively studied in geometry processing and physical simulation [17, 29, 49]. These approaches typically operate directly in *mesh space*, optimizing over vertex positions and leveraging penalty energies or interior-point formulations. While effective in simulation settings where a collision-free reference configuration is available, these methods are not naturally compatible with learning-based SMPL pipelines. In human pose estimation and motion generation, the optimization variable is the *pose parameter* $\boldsymbol{\theta}$ rather than raw mesh vertices. Moreover, a collision-free reference pose is generally unavailable. As a result, classical mesh-space solvers cannot be directly transferred to pose-space constrained optimization. In the domain of human body modeling, several works incorporate interpenetration penalties as "soft constraints" during model fitting or as auxiliary losses in learning-based predictors [4, 15, 19, 44, 57]. Recent volumetric human modeling approaches [36, 37] learn smooth articulated occupancy representations of posed human bodies. Such models provide continuous geometric fields that can implicitly encode collisions through occupancy evaluation. Other recent learning-based approaches [53, 54] train smooth neural collision classifiers and use them as differentiable constraints in post-hoc optimization. Unlike these methods, our approach is motivated by the regularity conditions required by gradient-based constrained optimization algorithms.

In this work, we formulate collision resolution as a constrained optimization problem that searches for the nearest collision-free pose to a self-intersecting configuration. Under this formulation, we introduce **PoseShield**, a neural collision constraint function defined directly in SMPL pose space, and solve the result-

ing problem using well-established gradient-based constrained optimization algorithms such as SLSQP and augmented Lagrangian methods [41]. For PoseShield to work reliably within these solvers, two requirements must be met. First, its sign must reliably indicate collision status: positive for collision-free poses and negative for self-intersecting ones. Second, these algorithms require the Linear Independence Constraint Qualification (LICQ) [41], which demands a nonvanishing constraint gradient away from the constraint boundary. Our key insight is the connection between this regularity requirement and the *Eikonal equation*: enforcing an Eikonal regularization on PoseShield ensures that its gradient norm is bounded away from zero throughout pose space, thereby satisfying LICQ and improving numerical stability. Such a constraint function is guaranteed to exist: the signed distance function to the collision boundary, characterized as the unique viscosity solution of the Eikonal equation [8]. Together, these properties yield a principled, self-contained post-hoc collision resolver for individual SMPL poses. We further extend our framework to human motion generation, enabling post-hoc self-collision correction for human motion sequences without knowing the generator. In summary, our contributions are:

- We propose a principled theoretical framework for self-collision resolution in SMPL pose space by formulating it as a constrained optimization problem with a learnable differentiable collision constraint. We further show that, under suitable assumptions on the learned collision field, gradient-based constrained solvers admit global and local convergence guarantees for this problem.
- We introduce **PoseShield**, a neural collision constraint trained with Eikonal regularization in pose space, and show how its training objective is designed to make these theoretical assumptions approximately hold in practice. In particular, we prove that the Eikonal training loss bounds the volume of pose-space regions where LICQ fails to hold, providing a quantitative connection between training accuracy and solver reliability.
- We demonstrate that the learned collision constraint can be seamlessly reused for temporally consistent motion correction without retraining the underlying motion generator, providing **a generator-agnostic post-processing module for human motion synthesis**.
- We perform evaluations on the newly curated **Human with Collisions** (HwC) dataset and the PROX dataset, where PoseShield substantially outperforms prior post-hoc collision-handling baselines.

2 Related Work

Human Self-Collision in SMPL-based Modeling. Self-collision is a long-standing issue for parametric human models like SMPL/SMPL-X, particularly under extreme articulations or depth ambiguities in monocular reconstruction. Early approaches primarily relied on coarse geometric heuristics, approximating the human body with primitive proxies such as capsules [4] to simplify intersection

queries. Subsequently, the field shifted toward mesh-level penetration penalties. For example, SMPLify-X [44] adapted differentiable self-interpenetration terms for expressive body capture, often implemented via distance-field-style losses [2, 57] and accelerated intersection detection [24]. This paradigm has been extended to scene-aware constraints in PROX [18]. For motion generation, recent work [19] employs efficient sphere-based proxies to incorporate self-intersection losses during the training of human motion models [16, 56]. Recently, implicit representations like COAP [36] and VolumetricSMPL [37] have advanced the state-of-the-art by demonstrating that learned occupancy fields can effectively reduce self-intersections through gradient-based refinement. While effective in reducing interpenetration for interaction modeling, these approaches are not explicitly designed to provide regularity guarantees for post-hoc constrained optimization. Relevantly, implicit neural representations have also been explored for modeling geometric or interaction constraints in the pose space [27]. Our work instead focuses on learning a collision constraint with properties tailored to support stable and theoretically grounded pose-space optimization.

General Mesh Collision Resolution. Classical collision-handling techniques, including penalty-based energies [5,6,13], interior-point formulations [11,17,21,29], and geometric frameworks like Repulsive Shells [49], share a fundamental limitation: they operate directly in mesh space by optimizing raw vertex positions. This makes them neither directly applicable nor computationally practical in pose-space optimization. Furthermore, these methods often necessitate a collision-free reference configuration, which is generally unavailable in modern pose-prediction pipelines [52, 53, 56, 64]. While specialized neural architectures have addressed collisions in specific domains, such as ContourCraft [14] and Self-Supervised Collision Handling [48] for garments, or Quaffure [51] for hair simulation, these solutions are tailored to their respective generative priors and do not generalize to the high-dimensional articulated manifold of human bodies. Ultimately, existing neural classifiers used for post-hoc resolution [52, 53, 61] often learn decision boundaries or proxy penetration depth signals for gradient guidance, but do not enforce global regularity conditions needed for robust constrained optimization.

Neural Solution of the Eikonal Equation. Solving the Eikonal equation [50, 66] subject to a sign constraint in 3D space corresponds to computing the Signed Distance Field (SDF) of an arbitrary shape. Thanks to its expressive power, the neural approximation of SDFs has become central to modern 3D generative models [30, 43, 60]. Although the SDF can represent highly complex geometry, these methods all assume a low-dimensional underlying domain—namely, the standard 3D Euclidean space. More recently, some research [39] demonstrated configuration-space distance fields whose gradients guide collision-free planning.

3 Problem: SMPL Self-Collision Resolution

In this section, we define the collision-resolution problem for the SMPL family of parametric human body models [33, 44]. An SMPL mesh is defined by shape

parameters $\boldsymbol{\beta} \in \mathbb{R}^{d_\beta}$ and pose parameters $\boldsymbol{\theta} \in \mathbb{R}^{d_\theta}$. Given $(\boldsymbol{\beta}, \boldsymbol{\theta})$, the SMPL function produces a mesh:

$$X = \mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta}),$$

where the mesh connectivity $\mathcal{T}$ is predefined and independent of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Due to imperfect motion capture or errors introduced by neural motion predictors [56], the generated SMPL meshes may exhibit self-collisions. In many practical scenarios, such as motion correction, the body shape remains fixed across frames. Therefore, we assume the shape parameter $\boldsymbol{\beta}$ is fixed and only optimize the pose parameter $\boldsymbol{\theta}$. We also ignore the global translation and rotation, since self-collision is invariant to them. To characterize whether a posed SMPL mesh is collision-free, we employ an exact mesh self-intersection test. Concretely, given a pose $\boldsymbol{\theta}$, we first decode it into a mesh $X = \mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta})$ and then apply a classical collision detector (e.g., FCL [42]) to obtain a binary collision indicator defined as:

$$\iota\big(\mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta})\big) \in \{-1, +1\},$$

which equals -1 if the mesh exhibits self-penetration and $+1$ otherwise. Given a self-colliding SMPL configuration $(\boldsymbol{\beta}, \boldsymbol{\theta}_0)$, our goal is to find a corrected pose $\boldsymbol{\theta}$ such that the decoded mesh:

$$X = \mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta}),$$

is collision-free while remaining visually close to the original configuration. Let $d_{\text{SMPL}}(\cdot, \cdot)$ denote a distance function that measures the discrepancy between two SMPL poses of a fixed shape. We formulate SMPL self-collision resolution as the following constrained optimization problem:

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} d_{\text{SMPL}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \quad \text{subject to} \quad \iota\big(\mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta})\big) = +1. \tag{1}$$

4 Learning a Differentiable Collision Constraint

As discussed in Sec. 3, the exact collision indicator $\iota(\cdot)$ provides a binary feasibility test. However, $\iota(\cdot)$ is non-differentiable with respect to the pose parameter $\boldsymbol{\theta}$, making gradient-based constrained optimization intractable. Modern constrained optimization algorithms, including Sequential Least Squares Programming (SLSQP), modified differential multiplier methods, and augmented Lagrangian approaches, require at least C^1 smooth constraint functions in order to compute well-defined gradients. Constraint functions must satisfy certain constraint qualifications such as LICQ for the iterates to reliably converge to a KKT point [3]. Since $\iota(\cdot)$ yields only binary feasibility labels and is discontinuous with zero gradient almost everywhere, these methods cannot be applied.

To enable robust constrained optimization in pose space, we introduce a surrogate, learnable neural constraint function $g(\boldsymbol{\theta})$, whose superlevel set $\{\boldsymbol{\theta} \mid g(\boldsymbol{\theta}) \geq 0\}$ approximates the set of collision-free SMPL poses under a fixed shape

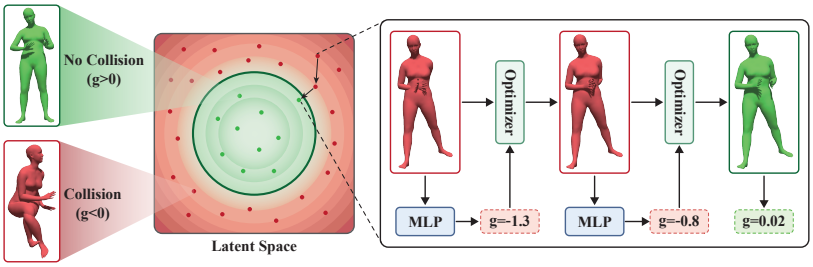


Fig. 2: Neural collision handling with PoseShield. PoseShield approximates the signed distance function (SDF) to the boundary between colliding and collision-free regions in the latent space. In practice, collision resolution algorithms optimize a self-penetrating sample toward its nearest point in the collision-free region, where the learned neural field provides local gradient guidance throughout the optimization process.

$\boldsymbol{\beta}$. With the surrogate function and given a self-colliding initial pose $\boldsymbol{\theta}_0$, the optimization problem becomes:

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} d_{\text{SMPL}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \quad \text{subject to} \quad g(\boldsymbol{\theta}) \geq C_l. \tag{2}$$

where C_l is a margin threshold. Theoretically, $C_l = 0$ is the optimal choice assuming a perfectly learned constraint g. In practice, however (as detailed in Sec. 5.2), we demonstrate that adjusting this threshold facilitates a controllable trade-off between maintaining geometric fidelity to the input pose and the effectiveness of collision resolution.

Our method is illustrated in Fig. 2. The neural collision-handling process operates in a latent space and follows a standard constrained optimization formulation, with g serving as a differentiable constraint function. We first analyze, in Sec. 4.1, the convergence behavior of standard gradient-based constrained solvers on this formulation under three assumptions on g. We then introduce in Sec. 4.2 our Eikonal regularization, which is designed to encourage the most non-trivial of these assumptions. A complete theoretical analysis is in the supplement. Building on these theoretical insights, Sec. 4.3 and Sec. 4.4 detail the practical training strategies for learning the differentiable collision constraint g. Finally, in Sec. 4.5, we generalize our framework from static pose optimization to accommodate continuous, temporally consistent human motion sequences.

4.1 Convergence Analysis

We analyze the convergence behavior of standard gradient-based constrained solvers on the surrogate problem in (2). Let $\Omega \subset [-1, 1]^{d_\theta}$ denote a bounded region of plausible SMPL poses in the 6D rotation representation. For the analysis, we set the threshold $C_l = 0$ and adopt the squared-Euclidean distance

$d_{\mathrm{SMPL}}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2$. We further assume the minimizer is interior to Ω and non-degenerate, so that $g \geq 0$ is the only active constraint.

Classical nonlinear optimization theory [41] establishes strong convergence guarantees for gradient-based constrained solvers under appropriate regularity conditions on the constraint function. We state three assumptions on the learned field g that together suffice for these guarantees in our setting.

Assumption 1 (Smoothness) *g is C^2 on Ω with Lipschitz-continuous gradient and Hessian.*

Assumption 2 (Feasibility consistency) *The sign of g correctly identifies the collision status: $g(\boldsymbol{\theta}) \geq 0$ if and only if $\boldsymbol{\theta}$ is a collision-free pose.*

Assumption 3 (Approximate Eikonal property) *There exists $\delta \in [0, 1)$ such that $1 - \delta \leq \|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta})\|_2 \leq 1 + \delta$ for all $\boldsymbol{\theta} \in \Omega$.*

The first assumption ensures that gradient-based methods are well-defined. The second ensures that the learned constraint faithfully represents the collision-free set. The third makes g behave like an approximate signed distance function in pose space; in particular, $\|\nabla g\| \geq 1 - \delta > 0$ implies that the Linear Independence Constraint Qualification (LICQ) holds globally. Under these three assumptions, standard gradient-based constrained solvers (e.g., SLSQP) admit both global and local convergence guarantees:

Theorem 1 (Global Convergence and Complexity). *Consider problem (2) under Assumptions 1, 2, and 3. Then:*

(i) ***Global LICQ:*** *The constraint qualification holds globally on Ω.*

(ii) ***Global Convergence:*** *From any starting pose $\boldsymbol{\theta}_0 \in \Omega$, a standard line-search SQP method with an ℓ_1 merit function (and sufficiently large penalty parameter) produces iterates whose every accumulation point is a first-order KKT point $(\boldsymbol{\theta}^\star, \lambda^\star)$.*

(iii) ***Iteration Complexity:*** *An ε-approximate KKT point—satisfying*

$$\left\|2(\boldsymbol{\theta}_k - \boldsymbol{\theta}_0) - \lambda_k \nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}_k)\right\| \leq \varepsilon, \quad |\min(0,\, g(\boldsymbol{\theta}_k))| \leq \varepsilon, \quad \lambda_k \geq 0, \quad |\lambda_k\, g(\boldsymbol{\theta}_k)| \leq \varepsilon,$$

is no harder to obtain than in unconstrained smooth optimization, whose worst-case first-order complexity is $\mathcal{O}(\varepsilon^{-2})$.

Theorem 2 (Local Convergence). *Consider problem (2) under Assumptions 1, 2, and 3. Let $\boldsymbol{\theta}^\star$ be a local minimizer, and assume the initial pose is infeasible, i.e., $g(\boldsymbol{\theta}_0) < 0$. Then:*

(i) ***LICQ and Strict Complementarity:*** *LICQ holds at $\boldsymbol{\theta}^\star$, and the unique KKT multiplier satisfies $0 < \frac{2}{1+\delta}\|\boldsymbol{\theta}^\star - \boldsymbol{\theta}_0\| \leq \lambda^\star \leq \frac{2}{1-\delta}\|\boldsymbol{\theta}^\star - \boldsymbol{\theta}_0\|$.*

(ii) ***SOSC and Fast Convergence:*** *Define $\kappa \triangleq \lambda^\star \|\nabla^2_{\boldsymbol{\theta}} g(\boldsymbol{\theta}^\star)\|_2$. If $\kappa < 2$, then the full Lagrangian Hessian is positive definite (implying Second-Order Sufficient Conditions), and SQP with exact Hessian converges locally to $(\boldsymbol{\theta}^\star, \lambda^\star)$ at a quadratic rate. A quasi-Newton (BFGS) variant satisfying the Dennis–Moré condition converges superlinearly.*

Proof. The complete proofs are provided in the supplementary material.

4.2 Eikonal Regularization

Among the three assumptions in Sec. 4.1, smoothness is automatically satisfied by an MLP with smooth activations (e.g., Softplus), and feasibility consistency is encouraged by direct sign supervision (Sec. 4.3). The approximate Eikonal property (Assumption 3), however, requires a dedicated regularizer. The ideal pointwise target is

$$\|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta})\| = 1, \tag{3}$$

which we encourage in expectation over Ω (assuming a uniform probability measure for analysis) by minimizing the average violation:

$$\begin{aligned} \mathcal{L}^i_{\text{grad}} &= \big|\|\nabla g(\boldsymbol{\theta}_i)\| - 1\big|, \\ \mathcal{L}_{\text{grad}} &= \mathbb{E}_{\boldsymbol{\theta}\sim\Omega}\Big[\big|\|\nabla g(\boldsymbol{\theta})\| - 1\big|\Big]. \end{aligned} \tag{4}$$

The following proposition shows that this loss provides a quantitative volume bound on the regions where Assumption 3 fails:

Proposition 1 (Volume Bound on Approximate Eikonal Failure). *Let S_δ denote the region where the approximate Eikonal condition fails for a given margin $\delta \in (0, 1)$:*

$$S_\delta = \left\{\boldsymbol{\theta} \in \Omega \;\middle|\; \big|\|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta})\| - 1\big| > \delta\right\}.$$

If the expected Eikonal loss satisfies $\mathcal{L}_{\text{grad}} \le \varepsilon$, then the probability measure of this failure region is bounded by:

$$\mathbb{P}(\boldsymbol{\theta} \in S_\delta) \le \frac{\varepsilon}{\delta}. \tag{5}$$

Proof. The complete proofs are provided in the supplementary material.

Combined with the fact that the approximate Eikonal property implies LICQ ($\|\nabla g\| \ge 1-\delta > 0$), this proposition provides a quantitative link between training accuracy and the volume of pose-space regions where LICQ is guaranteed to hold, directly supporting the convergence guarantees in Sec. 4.1.

4.3 Training Objective

To enforce Assumption 2, we impose boundary supervision using the exact collision indicator ι. Concretely, we encourage:

$$g(\boldsymbol{\theta})\, \iota(\mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta})) > 0,$$

so that $g(\boldsymbol{\theta}) > 0$ for collision-free poses and $g(\boldsymbol{\theta}) < 0$ otherwise. Together, Eikonal regularization and sign supervision encourage $g(\boldsymbol{\theta})$ to approximate a SDF-like function to the collision boundary in SMPL pose space. Using Monte

Carlo sampling, we construct: $\mathcal{D}_\theta = \{\langle \boldsymbol{\theta}_i, \iota_i \rangle\}_{i=1}^N$, where $\iota_i = \iota(\mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta}_i))$, and optimize the empirical PINNs-style [45] objective:

$$\begin{aligned} \mathcal{L}^i_{\text{sign}} &= -\min\big(g(\boldsymbol{\theta}_i)\iota_i, 0\big), \\ \mathcal{L}_{\text{Eikonal}} &= \frac{1}{|D_\theta|}\sum_{i=1}^{|D_\theta|}\left(\mathcal{L}^i_{\text{grad}} + \mathcal{L}^i_{\text{sign}}\right). \end{aligned} \tag{6}$$

This objective is a direct high-dimensional analogue of standard Eikonal training used in low-dimensional SDF learning [40, 43].

4.4 Temporal-Difference Variant

Inspired by the connection between $\mathcal{L}_{\text{grad}}$ and the Temporal Difference (TD) loss in offline reinforcement learning, and following Ni et al. [38] who show that a finite-timestep TD variant improves training, we replace $\mathcal{L}_{\text{grad}}$ with a symmetric TD loss parameterized by a small timestep Δt:

$$\mathcal{L}^i_{\text{TD}} = \big|g(\boldsymbol{\theta}_i + \boldsymbol{v}_i\Delta t) - g(\boldsymbol{\theta}_i - \boldsymbol{v}_i\Delta t) - 2\Delta t\big|, \tag{7}$$

where $\boldsymbol{v}_i = \nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}_i)/\|\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}_i)\|$ is the normalized velocity function. We find that replacing $\mathcal{L}_{\text{grad}}$ with $\mathcal{L}_{\text{TD}}$ improves the performance as demonstrated in Tab. 1.

4.5 Collision Resolution for Motion Sequence

Our formulation naturally extends from static pose optimization to human motion sequences. Specifically, the sampling process in human motion diffusion models [9, 20, 28, 31, 35, 47, 56, 59, 62, 67] and human motion flow matching models [55] can be formulated as a differentiable generative mapping: $\mathbf{m} = f(\mathbf{x})$, where $\mathbf{x}$ denotes the input noise, and $\mathbf{m}$ is the generated human motion sequence. To enforce task-specific constraints (e.g., obstacle or self-collision avoidance), DNO [25] proposes a conditional motion synthesis formulated as:

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \mathcal{Q}\big(f(\mathbf{x})\big), \tag{8}$$

where $\mathcal{Q}$ is a user-defined criterion measuring constraint satisfaction or motion plausibility. This formulation is naturally compatible with our constrained optimization formulation in Equation 2. Specifically, we first train PoseShield as usual for static human poses. Suppose we are given a motion sequence with self-collisions defined as $\mathbf{m}_s = [\boldsymbol{\theta}^0_s, \boldsymbol{\theta}^1_s, \cdots, \boldsymbol{\theta}^T_s]$, and let the optimization variable be $\mathbf{m} = [\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \cdots, \boldsymbol{\theta}^T]$, we define the objective function $\mathcal{Q}$ as follows:

$$\mathcal{Q}(\mathbf{m}) = \sum_{i=0}^{T} \max\big(C_l - g(\boldsymbol{\theta}^i), 0\big) + \lambda_m\, d_{motion}(\mathbf{m}, \mathbf{m}_s), \tag{9}$$

where the first term penalizes violations of the collision constraint, and the second term enforces proximity to the original motion. The coefficient λ_m balances

collision resolution and motion preservation. The distance metric d_{motion} is detailed in the supplementary material. Note that we avoid the hard constraints from Equation 2, which are disallowed by DNO, and replace them with soft constraint functions.

5 Evaluation

In this section, we evaluate the performance of PoseShield and compare it against baselines. We first discuss collision resolution for static human poses (Section 5.1), and then ablate key design choices and demonstrate properties of our method (Section 5.2). Next, we scale the constraint function learned from individual poses to human motion sequences (Section 5.3).

5.1 Application: Static Human Pose

Humans with Collisions (HwC) Dataset. Despite the availability of various human pose datasets [10, 22, 34], meshes sampled solely from the collision-free ground-truth distribution do not expose PoseShield to any self-colliding examples. To address this, we introduce the **Humans with Collisions (HwC)** dataset of nearly $931k$ SMPL poses, which is detailed in the supplementary. A subset of 500 representative self-penetrating meshes serves for benchmarking. This dataset also serves as the sampling pool to approximate Ω in Sec. 4.2. In addition, we follow previous work [36] and evaluate the methods on a subset of PROX dataset [18].

Metrics. Following [53], we adopt the following metrics:

- Success Rate (SCC): The rate of the collision resolution method producing penetration-free meshes.
- Penetration Depth Reduction (PDR): The ratio of reduced penetration depth (PD) [42] relative to the original penetration depth defined as:

$$\text{PDR} = \max\left[1 - \frac{\text{PD after optimization}}{\text{PD before optimization}}, 0\right].$$

- Mean Vertex Distance (MVD): The average L_2 distance of the vertices between the original and the optimized mesh. (An ideal collision handler resolves collisions while keeping the resulting pose as close as possible to the original sample.)

Baselines. We consider the following baselines: 1) Torch-mesh-isect [57]: An open-source tool developed specifically for SMPL-based collision resolution. 2) Classifier-baseline: Inspired by N-Penetrate [53], we replace the collision constraint function with the probability of being a non-colliding mesh predicted by a classifier using the cross-entropy loss. More details are deferred to the supplementary. 3) COAP [36]: a volumetric occupancy-field approach that treats

Table 1: Quantitative results and ablation study on the pose datasets. Our method achieves significant improvement in collision resolution ability over all baselines. ↑: higher values are better; ↓: lower values are better. Bold and underlined values indicate the **best** and second-best performers in each category, respectively. $\mathcal{L}_{\text{grad}}$ and $\mathcal{L}_{\text{TD}}$ represent the gradient loss terms; WD denotes the weighted distance metric in pose space (see Sec. 5.2).

Method	$\mathcal{L}_{\text{grad}}$	$\mathcal{L}_{\text{TD}}$	WD	HwC			PROX [18]		
				SCC↑	PDR↑	MVD↓	SCC↑	PDR↑	MVD↓
Torch-mesh-isect [57]				0.100	0.357	0.041	0.110	0.291	0.012
Classifier baseline				0.056	0.081	**0.002**	0.170	0.204	**0.006**
COAP [36]				0.446	0.832	0.106	0.560	0.775	0.016
VolumetricSMPL [37]				0.250	0.541	0.068	0.333	0.699	0.013
Ours		✓	✓	0.958	0.982	0.059	0.800	0.893	0.021
Ours (w/o WD)		✓		**0.960**	**0.987**	0.067	**0.850**	**0.918**	0.036
Ours ($\mathcal{L}_{\text{grad}}$)	✓		✓	0.862	0.917	0.062	0.520	0.615	0.019
Ours ($\mathcal{L}_{\text{grad}}$+$\mathcal{L}_{\text{TD}}$)	✓	✓	✓	0.870	0.922	0.062	0.670	0.743	0.022
Ours (w/o grad term)			✓	0.068	0.081	0.458	0.050	0.106	0.431

collision resolution as a sampling-based occupancy penalty in the 3D workspace rather than a direct pose-space constraint. 4) VolumetricSMPL [37]: a follow-up volumetric body representation that extends COAP by modeling the human body as a signed distance field (SDF) instead of an occupancy field.

Implementation Details. For optimization, we use the standard SLSQP method implemented in SciPy [58]. The network is a 12-layer MLP with a hidden dimension of 512. We train it for 200 epochs, where we adopt active learning [53] to collect boundary samples every 40 epochs. The entire training process requires around 17 hours on a single GPU. For the loss term, the default Δt in Equation 7 is 0.01. Inference takes 7.26 seconds per pose on average.

Quantitative Results. As shown in Tab. 1, our method substantially outperforms all baselines in both success rate and penetration depth reduction. The classifier baseline, which also performs latent-space optimization, achieves a very low success rate, highlighting that our proposed neural field is a crucial component for effective collision handling within such frameworks. Similarly, Torch-mesh-isect [57] applies triangle-level penetration losses in pose optimization, but its local face-level loss formulation [44] prevents it from resolving deep self-penetrations, often causing it to converge to local minima and fail to move the mesh sufficiently. Among the baselines, COAP [36] attains the highest SCC and PDR, but its performance degrades in challenging cases with severe penetrations. Most notably, our method increases the success rate on the HwC dataset from 0.446 to 0.958 while achieving a significantly lower MVD, which is also reflected in Fig. 5. This indicates that our formulation resolves self-collisions with

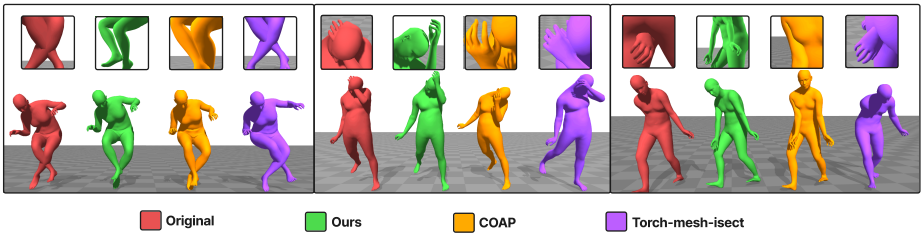


Fig. 3: Qualitative comparison with baseline methods on three cases (left to right). Within each case, results are shown from left to right: Original input, Ours, COAP [36], and Torch-mesh-isect [57]. Our method consistently removes self-collisions. Insets highlight representative local self-collision regions. Torch-mesh-isect fails to resolve the collisions in all three cases, while COAP almost resolves the first case but still leaves minor residual intersections.

smaller pose deviations, empirically approximating minimal-distance corrections to collision-free configurations.

Qualitative Results. Fig. 3 presents qualitative comparisons with baseline collision-handling methods. Across the three examples, our method consistently removes the highlighted self-intersections, producing visually collision-free configurations in the inset regions. In contrast, Torch-mesh-isect fails to resolve the collisions in all three cases, which is consistent with its low success rate reported in Tab. 1. COAP is able to reduce penetrations in some cases and can nearly eliminate the collision in relatively simple scenarios (e.g., the first case), although small residual intersections often remain. This behavior is also reflected in Tab. 1, where COAP achieves relatively high PDR but still falls short of fully resolving collisions in some cases. In more complex cases involving multiple body contacts (e.g., the second and third cases), COAP is unable to remove the intersections. Overall, our method achieves more reliable collision resolution while preserving the overall pose structure of the input.

5.2 Ablation Study

Choices of $d_{\text{SMPL}}(\boldsymbol{\theta}, \boldsymbol{\theta}')$. We ablate the choice of the pose distance metric. We compare the standard L_2 distance:

$$d_{\text{std}}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{J} \sum_{j=1}^{J} ||\boldsymbol{\theta}_j - \boldsymbol{\theta}'_j||_2, \tag{10}$$

against a weighted L_2 distance:

$$d_{\text{WD}}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{\sum_{j=1}^{J} w_j} \sum_{j=1}^{J} w_j ||\boldsymbol{\theta}_j - \boldsymbol{\theta}'_j||_2, \tag{11}$$

where $\boldsymbol{\theta}_j \in \mathbb{R}^6$ denotes the 6D parameters of the j-th joint, and w_j represents the weight assigned based on the size of the subtree rooted at joint j within

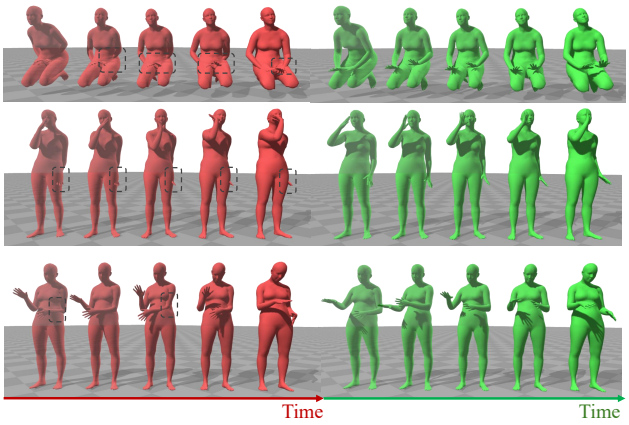


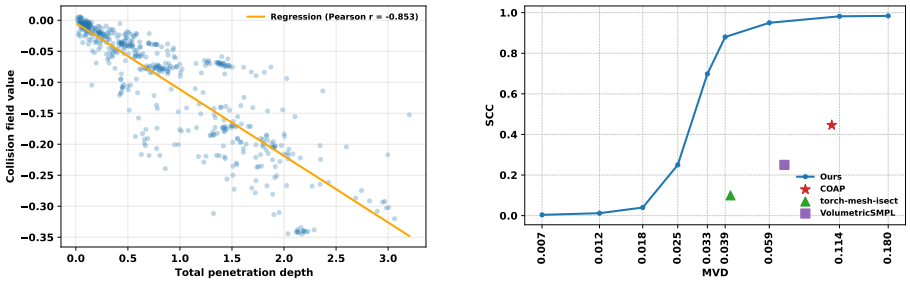
Fig. 4: Our method resolves the collision in the original samples while preserving the overall motion structure. The red ones are original samples, and the green ones are optimized ones.

the kinematic hierarchy. We refer to the former as "Ours w/o WD" (Weighted Distance). As shown in Tab. 1, while "Ours w/o WD" achieves slightly higher collision resolution rates, it suffers from a higher Mean Vertex Distance (MVD). In contrast, our full model with WD significantly reduces MVD from 0.067 to 0.059. This reduction is achieved by penalizing rotations of proximal joints more heavily, as their perturbations propagate through the kinematic hierarchy and cause large-scale displacements of downstream subtrees.

$\mathcal{L}_{TD}$ is sufficient. As shown in Tab. 1, using $\mathcal{L}_{TD}$ alone yields the best overall performance across all metrics. We conjecture that incorporating $\mathcal{L}_{grad}$ introduces second-order derivatives during optimization, which increases training instability and hinders convergence. In contrast, $\mathcal{L}_{TD}$ effectively enforces the local geometric consistency. The variant that omits both loss terms fails to approximate a solution to the Eikonal equation, leading to poor performance.

Tradeoff between SCC and MVD. As shown in Fig. 5b, increasing the constraint margin C_l in Equation 2 leads to higher SCC while also increasing MVD. This reflects the inherent trade-off between collision resolution and geometric fidelity, as also observed in [53]. Importantly, our method provides a controllable mechanism to navigate this trade-off simply by tuning C_l. As illustrated in Fig. 5b, our approach consistently achieves higher SCC at the same or lower level of MVD compared with baseline methods, indicating a more favorable trade-off between collision removal and motion preservation.

Correlation between g and penetration depth. Our method can learn the degree of collision, even though the dataset provides only binary collision labels. While this latent function is not directly visualizable, Fig. 5a shows that its values are correlated with the total penetration depth on the training set.



(a) Correlation between g values and PD. **(b)** Trade-off between SCC and MVD.

Fig. 5: Analysis of the neural collision constraint. (a) Our neural field $g(\boldsymbol{\theta})$ shows a strong correlation with physical penetration depth (PD). We sample 500 instances for visualization. (b) The threshold C_l provides a controllable trade-off between collision resolution success (SCC) and geometric fidelity (MVD). The points correspond to $C_l \in \{-0.2, -0.1, -0.05, 0, 0.05, 0.1, 0.2, 0.4, 0.6\}$, ordered left to right. Performance from baseline methods is also included for reference.

5.3 Application: Human Motion Sequence

We obtain the trained g from Sec. 5.1 and show that g can be robustly generalized to human motion sequences. We utilize a pre-trained motion model [55] as the generative prior f and select motion sequences with self-penetration from a motion dataset [1]. Examples are presented in Fig. 4. Our method effectively resolves collisions while preserving the overall motion and avoiding noticeable artifacts. More quantitative results are provided in the supplementary material.

6 Conclusion

We presented PoseShield, a neural collision constraint defined directly in SMPL pose space for post-hoc self-collision resolution. By establishing a connection between collision handling and the Eikonal equation, we provide theoretical grounding for neural constraint learning. In particular, we showed that Eikonal-regularized constraint functions satisfy the LICQ, ensuring the feasibility and numerical stability of constrained optimization in the pose space. The same learned constraint further serves as a generator-agnostic post-hoc corrector for human motion sequences, requiring no retraining of the underlying motion model. Experiments validate that PoseShield significantly improves collision resolution success rates compared to prior state-of-the-art approaches.

References

1. Athanasiou, N., Cseke, A., Diomataris, M., Black, M.J., Varol, G.: Motionfix: Text-driven 3d human motion editing. In: SIGGRAPH Asia 2024 Conference Papers. pp. 1–11 (2024)

2. Ballan, L., Taneja, A., Gall, J., Van Gool, L., Pollefeys, M.: Motion capture of hands in action using discriminative salient points. In: European Conference on Computer Vision. pp. 640–653. Springer (2012)
3. Bertsekas, D.P.: Nonlinear programming. Journal of the Operational Research Society **48**(3), 334–334 (1997)
4. Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., Black, M.J.: Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14. pp. 561–578. Springer (2016)
5. Chen, A.H., Hsu, J., Liu, Z., Macklin, M., Yang, Y., Yuksel, C.: Offset geometric contact. ACM Transactions on Graphics (TOG) **44**(4), 1–21 (2025)
6. Chen, H., Diaz, E., Yuksel, C.: Shortest path to boundary for self-intersecting meshes. ACM Transactions on Graphics (TOG) **42**(4), 1–15 (2023)
7. Choutas, V., Müller, L., Huang, C.H.P., Tang, S., Tzionas, D., Black, M.J.: Accurate 3d body shape regression using metric and semantic attributes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2718–2728 (2022)
8. Crandall, M.G., Ishii, H., Lions, P.L.: User's guide to viscosity solutions of second order partial differential equations. Bulletin of the American mathematical society **27**(1), 1–67 (1992)
9. Dai, W., Chen, L.H., Wang, J., Liu, J., Dai, B., Tang, Y.: Motionlcm: Real-time controllable motion generation via latent consistency model. In: European Conference on Computer Vision. pp. 390–408. Springer (2024)
10. Delmas, G., Weinzaepfel, P., Lucas, T., Moreno-Noguer, F., Rogez, G.: Posescript: Linking 3d human poses and natural language. IEEE transactions on pattern analysis and machine intelligence (2024)
11. Fang, Y., Li, M., Jiang, C., Kaufman, D.M.: Guaranteed globally injective 3d deformation processing. ACM Transactions on Graphics **40**(4) (2021)
12. Feng, Y., Lin, J., Dwivedi, S.K., Sun, Y., Patel, P., Black, M.J.: Chatpose: Chatting about 3d human pose. In: CVPR (2024)
13. Fisher, S., Lin, M.C.: Deformed distance fields for simulation of non-penetrating flexible bodies. In: Computer Animation and Simulation 2001: Proceedings of the Eurographics Workshop in Manchester, UK, September 2–3, 2001. pp. 99–111. Springer (2001)
14. Grigorev, A., Becherini, G., Black, M., Hilliges, O., Thomaszewski, B.: Contourcraft: Learning to resolve intersections in neural multi-garment simulations. In: ACM SIGGRAPH 2024 conference papers. pp. 1–10 (2024)
15. Guan, P., Weiss, A., Balan, A.O., Black, M.J.: Estimating human shape and pose from a single image. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 1381–1388. IEEE (2009)
16. Guo, C., Mu, Y., Javed, M.G., Wang, S., Cheng, L.: Momask: Generative masked modeling of 3d human motions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1900–1910 (2024)
17. Harmon, D., Vouga, E., Smith, B., Tamstorf, R., Grinspun, E.: Asynchronous contact mechanics. ACM Trans. Graph. **28**(3) (Jul 2009)
18. Hassan, M., Choutas, V., Tzionas, D., Black, M.J.: Resolving 3D human pose ambiguities with 3D scene constraints. In: International Conference on Computer Vision. pp. 2282–2292 (Oct 2019)

19. Herrmann, P., Bieshaar, M., Mack, D., Herzog, P.R., Gall, J.: Self-intersection-aware 3d human motion generation using an efficient human sphere proxy. In: British Machine Vision Conference (BMVC) (2025), poster session, Wednesday 26th November, Paper No. 737
20. Hong, S., Kim, C., Yoon, S., Nam, J., Cha, S., Noh, J.: Salad: Skeleton-aware latent diffusion for text-driven motion generation and editing. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 7158–7168 (2025)
21. Huang, Z., Araújo, C., Kunz, A., Zorin, D., Panozzo, D., Zordan, V.: Intersection-free garment retargeting. In: Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers. pp. 1–12 (2025)
22. Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE transactions on pattern analysis and machine intelligence **36**(7), 1325–1339 (2013)
23. Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J.: End-to-end recovery of human shape and pose. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7122–7131 (2018)
24. Karras, T.: Maximizing parallelism in the construction of bvhs, octrees, and k-d trees. In: Proceedings of the Fourth ACM SIGGRAPH / Eurographics Conference on High-Performance Graphics. pp. 33–37. Eurographics Association (2012)
25. Karunratanakul, K., Preechakul, K., Aksan, E., Beeler, T., Suwajanakorn, S., Tang, S.: Optimizing diffusion noise can serve as universal motion priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1334–1345 (2024)
26. Kocabas, M., Athanasiou, N., Black, M.J.: Vibe: Video inference for human body pose and shape estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5253–5263 (2020)
27. Kulkarni, N., Rempe, D., Genova, K., Kundu, A., Johnson, J., Fouhey, D., Guibas, L.: Nifty: Neural object interaction fields for guided human motion synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 947–957 (2024)
28. Li, J., Wu, J., Liu, C.K.: Object motion guided human motion synthesis. ACM Transactions on Graphics (TOG) **42**(6), 1–11 (2023)
29. Li, M., Kaufman, D.M., Jiang, C.: Codimensional incremental potential contact. ACM Trans. Graph. (SIGGRAPH) **40**(4) (2021)
30. Li, M., Duan, Y., Zhou, J., Lu, J.: Diffusion-sdf: Text-to-shape via voxelized diffusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12642–12651 (2023)
31. Li, Z., Cheng, K., Ghosh, A., Bhattacharya, U., Gui, L., Bera, A.: Simmotionedit: Text-based human motion editing with motion similarity prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2025)
32. Lin, J., Zeng, A., Lu, S., Cai, Y., Zhang, R., Wang, H., Zhang, L.: Motion-x: A large-scale 3d expressive whole-body human motion dataset. Advances in Neural Information Processing Systems **36**, 25268–25280 (2023)
33. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. ACM Transactions on Graphics **34**(6), 248:1–248:16 (2015)

34. Mahmood, N., Ghorbani, N., Troje, N.F., Pons-Moll, G., Black, M.J.: Amass: Archive of motion capture as surface shapes. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5442–5451 (2019)
35. Meng, Z., Xie, Y., Peng, X., Han, Z., Jiang, H.: Rethinking diffusion for text-driven human motion generation: Redundant representations, evaluation, and masked autoregression. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 27859–27871 (2025)
36. Mihajlovic, M., Saito, S., Bansal, A., Zollhoefer, M., Tang, S.: Coap: Compositional articulated occupancy of people. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13201–13210 (2022)
37. Mihajlovic, M., Zhang, S., Li, G., Zhao, K., Muller, L., Tang, S.: Volumetricsmpl: A neural volumetric body model for efficient interactions, contacts, and collisions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5060–5070 (2025)
38. Ni, R., zherong pan, Qureshi, A.H.: Physics-informed temporal difference metric learning for robot motion planning. In: The Thirteenth International Conference on Learning Representations (2025)
39. Ni, R., Qureshi, A.H.: NTFields: Neural time fields for physics-informed robot motion planning. In: The Eleventh International Conference on Learning Representations (2023)
40. Ni, R., Schneider, T., Panozzo, D., Pan, Z., Gao, X.: Robust & asymptotically locally optimal uav-trajectory generation based on spline subdivision. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 7715–7721. IEEE (2021)
41. Nocedal, J., Wright, S.J.: Numerical optimization. Springer (2006)
42. Pan, J., Chitta, S., Manocha, D.: Fcl: A general purpose library for collision and proximity queries. In: 2012 IEEE international conference on robotics and automation. pp. 3859–3866. IEEE (2012)
43. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 165–174 (2019)
44. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A.A., Tzionas, D., Black, M.J.: Expressive body capture: 3d hands, face, and body from a single image. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2019)
45. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational physics **378**, 686–707 (2019)
46. Romero, J., Tzionas, D., Black, M.J.: Embodied hands: Modeling and capturing hands and bodies together. arXiv preprint arXiv:2201.02610 (2022)
47. Ruiz-Ponce, P., Barquero, G., Palmero, C., Escalera, S., García-Rodríguez, J.: Mixermdm: Learnable composition of human motion diffusion models. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 12380–12390 (2025)
48. Santesteban, I., Thuerey, N., Otaduy, M.A., Casas, D.: Self-supervised collision handling via generative 3d garment models for virtual try-on. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11763–11773 (2021)

49. Sassen, J., Schumacher, H., Rumpf, M., Crane, K.: Repulsive shells. ACM Transactions on Graphics **43**(4), 1–22 (2024)
50. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. proceedings of the National Academy of Sciences **93**(4), 1591–1595 (1996)
51. Stuyck, T., Lin, G.W.C., Larionov, E., Chen, H.y., Bozic, A., Sarafianos, N., Roble, D.: Quaffure: Real-time quasi-static neural hair simulation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 239–249 (2025)
52. Tan, Q., Pan, Z., Manocha, D.: Lcollision: Fast generation of collision-free human poses using learned non-penetration constraints. In: Proceedings of the AAAI Conference on Artificial Intelligence (2021)
53. Tan, Q., Pan, Z., Smith, B., Shiratori, T., Manocha, D.: N-penetrate: Active learning of neural collision handler for complex 3d mesh deformations. In: ICML. pp. 21037–21049 (2022)
54. Tan, Q., Zhou, Y., Wang, T., Ceylan, D., Sun, X., Manocha, D.: A repulsive force unit for garment collision handling in neural networks. In: European conference on computer vision. pp. 451–467. Springer (2022)
55. Team, T.H.D.D.H.: Hy-motion 1.0: Scaling flow matching models for text-to-motion generation. arXiv preprint arXiv:2512.23464 (2025)
56. Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-or, D., Bermano, A.H.: Human motion diffusion model. In: The Eleventh International Conference on Learning Representations (2022)
57. Tzionas, D., Ballan, L., Srikantha, A., Aponte, P., Pollefeys, M., Gall, J.: Capturing hands in action using discriminative salient points and physics simulation. International Journal of Computer Vision **118**(2), 172–193 (2016)
58. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods **17**, 261–272 (2020)
59. Xu, S., Li, Z., Wang, Y.X., Gui, L.Y.: Interdiff: Generating 3d human-object interactions with physics-informed diffusion. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14928–14940 (2023)
60. Yariv, L., Puny, O., Gafni, O., Lipman, Y.: Mosaic-sdf for 3d generative models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4630–4639 (2024)
61. Zesch, R.S., Modi, V., Sueda, S., Levin, D.I.: Neural collision fields for triangle primitives. In: SIGGRAPH Asia 2023 Conference Papers. pp. 1–10 (2023)
62. Zhang, M., Guo, X., Pan, L., Cai, Z., Hong, F., Li, H., Yang, L., Liu, Z.: Remodiffuse: Retrieval-augmented motion diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 364–373 (2023)
63. Zhang, M., Jin, D., Gu, C., Hong, F., Cai, Z., Huang, J., Zhang, C., Guo, X., Yang, L., He, Y., et al.: Large motion model for unified multi-modal motion generation. In: European Conference on Computer Vision. pp. 397–421. Springer (2024)
64. Zhang, S., Bhatnagar, B.L., Xu, Y., Winkler, A., Kadlecek, P., Tang, S., Bogo, F.: Rohm: Robust human motion reconstruction via diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14606–14617 (2024)

65. Zhang, T., Huang, B., Wang, Y.: Object-occluded human shape and pose estimation from a single color image. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7376–7385 (2020)
66. Zhao, H.: A fast sweeping method for eikonal equations. Mathematics of computation **74**(250), 603–627 (2005)
67. Zhong, L., Xie, Y., Jampani, V., Sun, D., Jiang, H.: Smoodi: Stylized motion diffusion model. In: European Conference on Computer Vision. pp. 405–421. Springer (2024)